

**PATENT APPLICATION**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of

Marc DYMETMAN et al.

Group Art Unit: 2626

Application No.: 10/727,614

Examiner: J. RIDER

Filed: December 5, 2003

Docket No.: 115798

For: SYSTEMS AND METHODS FOR SEMANTIC STENOGRAPHY

**DECLARATION UNDER 37 C.F.R. §1.131**

We, Marc Dymetman, Caroline Brun and Aurélien Max, citizens of France, hereby declare and state:

1. This Declaration is submitted as evidence that the subject matter claimed in the above-identified application was invented prior to April, 2003, the publication date of Max et al., "Reversing Controlled Document Authoring to Normalize Documents," in the proceeding of EACL '03 Student Research Workshop, Budapest, Hungary, 2003, pp. 33-40 ("Max").

2. We are the named inventors of the above-identified application.

3. We are also the co-authors of the attached Invention Proposal ("IP") dated prior to April, 2003, a true copy of which appears as Exhibit A attached to this Declaration.

4. In the copy of the IP attached hereto as Exhibit A, dates and other material which could indicate dates have been masked out. Additionally, the employee numbers on page 1 have been masked out, as have all references to internal proprietary Xerox Corporation research and development programs.

5. Exhibit A describes a method for stenographically processing input data, including, receiving short note input data, using a semantic grammar to generate semantic structure, producing with a first realization grammar a plurality of local text realizations from the semantic structure, matching the short note input data with ones of the plurality of local text realizations to define a final semantic structure, producing with a second realization grammar global text realizations from the final semantic structure.

Exhibit A also describes a system for stenographically processing input data, including an input device which receives short note input data, a semantic grammar generator which uses a semantic grammar to generate semantic structure, a local text realization generator which produces with a first realization grammar a plurality of local text realizations from the semantic structure, a processor that matches the short note input data with the plurality of local text realizations to define a final semantic structure, and the processor that produces with a second realization grammar global text realizations from the final semantic structure.

Exhibit A further describes a computer program product, including a computer usable medium having computer readable program code embodied therein for converting input data into a global text realization, wherein said computer readable instructions includes a computer readable program code for causing a computer to receive input data, a computer readable program code for causing the computer to generate a global text realization based on the input data, and a computer readable program code for causing a computer to output the global text realization.

Exhibit A still further describes a computer program product, including a computer usable medium having computer readable program code embodied therein for converting short notes into a global text realization, wherein said computer readable instructions including a computer readable program code for causing a computer to perform a fuzzy match

between local text realizations and short notes to provide at least one local text realization in association with each short note, and a computer readable program code for causing the computer to generate a global text realization for each short note from associated local text realization selected by an operator.

Exhibit A even further describes a system for converting short notes into a global text realization including means for inputting short notes, means for generating a global text realization based on the short notes, and means for outputting the global text realization.

In particular, the specifics of the stenographically processing input data method and system disclosed in the above-identified application are described on pages 6 through 17 of Exhibit A.

6. The invention described in Exhibit A may thus be summarized as follows:

(a) a method for stenographically processing input data, including, receiving short note input data, using a semantic grammar to generate semantic structure, producing with a first realization grammar a plurality of local text realizations from the semantic structure, matching the short note input data with ones of the plurality of local text realizations to define a final semantic structure, producing with a second realization grammar global text realizations from the final semantic structure;

(b) a system for stenographically processing input data, including an input device which receives short note input data, a semantic grammar generator which uses a semantic grammar to generate semantic structure, a local text realization generator which produces with a first realization grammar a plurality of local text realizations from the semantic structure, a processor that matches the short note input data with the plurality of local text realizations to define a final semantic structure, and the processor that produces with a second realization grammar global text realizations from the final semantic structure;

(c) a computer program product, including a computer usable medium having computer readable program code embodied therein for converting input data into a global text realization, wherein said computer readable instructions includes a computer readable program code for causing a computer to receive input data, a computer readable program code for causing the computer to generate a global text realization based on the input data, and a computer readable program code for causing a computer to output the global text realization;

(d) a computer program product, including a computer usable medium having computer readable program code embodied therein for converting short notes into a global text realization, wherein said computer readable instructions including a computer readable program code for causing a computer to perform a fuzzy match between local text realizations and short notes to provide at least one local text realization in association with each short note, and a computer readable program code for causing the computer to generate a global text realization for each short note from associated local text realization selected by an operator; and

(e) a system for converting short notes into a global text realization including means for inputting short notes, means for generating a global text realization based on the short notes, and means for outputting the global text realization.

7. Exhibit A describes an invention conceived and reduced to practice prior to April, 2003. This invention is claimed in the above-identified application.

8. Prior to April, 2003, we and/or those under our control and supervision, carried out a reduction to practice of the invention described in Exhibit A and thereby provided a stenographically processing input data method and system as described in paragraphs 5-7 herein.

Application No. 10/430,311

9. We hereby declare that all statements made herein of our own knowledge are true, and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine and/or imprisonment under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing therefrom.

Date: 12 october 2007

  
\_\_\_\_\_  
Marc Dymetman

Date: 15 october 2007

  
\_\_\_\_\_  
Caroline Brun

Date: \_\_\_\_\_

\_\_\_\_\_  
Aurélien Max

Application No. 10/430,311

we're made with the knowledge that willful false statements and the like  
so made are punishable by fine and/or imprisonment under Section 1001  
of Title 18 of the United States Code, and that such willful false  
statements may jeopardize the validity of the application or any patent  
issuing therefrom.

Date

:

Marc Dymetman

Date

:

Caroline Brun

Date

:

October 16, 2007

A. Max

Aurélien Max

# XEROX Invention Proposal **RECEIVED**

(For Patent Dept. use only)

Please TYPE and return, accompanied by (1) Manager's Comments form and (2) Inventorship Statement form(s) to: Val Whitelaw, Patent Department, XLTC, Welwyn Garden City, Herts AL7 1HE, UK. (Val.Whitelaw@GBR.XEROX.COM)

IP No:

Atty:

XPC:

**XEROX CONFIDENTIAL**

**INTELLECTUAL PROPERTY  
DEPARTMENT**

Proposal submitted by: (If space for additional submitters is required, please use an additional sheet)

1	Name: Marc Dymetman	Internal Tel. No:
	Internal Address XRCE	Email Address:
2	Name: Caroline Brun	Internal Tel. No:
	Internal Address: XRCE	Email Address:
3	Name: Aurélien Max	Internal Tel. No:
	Internal Address: XRCE	Email Address:

Manager: Pierre Isabelle Internal Address: XRCE Internal Tel. No:

Title of invention: *Semantic Stenography*

Name of Program, Product or Technology: Content-Analysis, XRCE

Name of others known to have done similar work: see prior art section in the invention description

List any similar or related Invention Proposals, patents, publications or products:  
see prior art section in the invention description

Indicate the date of any previous or planned future disclosure of the invention external to Xerox and describe the nature of the disclosure: No publication planned at this point, but this may change

Any outside funding and/or contractual relationships connected with the work described herein: none

Are any of the inventors non-Xerox employees? NO

Extent of implementation:

- |                                  |                      |
|----------------------------------|----------------------|
| a) Paper proposal Yes            | c) Prototype         |
| b) Feasibility model/calculation | d) Production design |

Provide a brief summary or abstract of the invention, specifically pointing out the features or application you think are new or beneficial:

We propose a method allowing writers to jot down a set of short notes consisting in **semantic abbreviations** for complex concepts, in a restricted domain of discourse. These short notes are then automatically converted into a **semantically coherent grammatical text** which reflects the content of the notes. A **user interface** is also provided permitting the user to inspect the meanings attributed by the system to the short notes, and to make corrections to these meanings if necessary.

The invention disclosure has three aspects, in decreasing order of **novelty significance**:

1. A **novel paradigm for producing text: semantic stenography**. This is the most novel and important aspect of the invention.
2. A **proposed technical embodiment** (method) for this paradigm. This is also important, but other embodiments of the invention are possible; this aspect of the invention has partial intersection with prior art, such as Aurélien's Max publications about his PhD work at XRCE, but also **some novel aspects** (short notes versus full input document, user interface, simplification of search procedure).
3. Disclosure of an **application** domain: job offer announcements. This domain has been chosen mainly for illustration purposes and is only one among many possible domains (commercial correspondence, classified ads, CRM through email, doctors' referral letters, mobile telephone interfaces, ...)

Submitter(s) Signature(s)

Date:

Witnessed & understood by:

Date:

PRINT NAME:

I/We have also completed an Inventorship Statement Form

**XEROX CONFIDENTIAL**

Version

# XEROX Invention Proposal Form

Please TYPE and return, accompanied by (1) Manager's Comments form and (2) Inventorship Statement form(s) to: Val Whitelaw, Patent Department, XLTC, Welwyn Garden City, Herts AL7 1HE, UK. ([Val.Whitelaw@GBR.XEROX.COM](mailto:Val.Whitelaw@GBR.XEROX.COM))

(For Patent Dept. use only)

IP No:

Atty:

XPC:

**XEROX CONFIDENTIAL**

**Description of the invention** - (This should include: 1) an explanation of the problem solved by the invention; 2) description of how the invention works - with drawings, where possible; and 3) a discussion of how the invention improves over present technology. It would also be helpful if you could say whether there are alternatives available. If so, what are the relative advantages of the present proposal?)

See enclosed document.

(Please continue - if filling in this form electronically further pages will automatically be generated as required.)

Submitter(s) Signature(s)

Date:

Witnessed & understood by:

Date:

PRINT NAME:

I/We have also completed an Inventorship Statement Form

**XEROX CONFIDENTIAL**

Version



**XEROX****Manager's Comments Form**

(This form to be TYPED and attached to Invention Proposal Form and Inventorship Statement Form)

(For Patent Dept. use only)

IP No: [REDACTED]

**XEROX CONFIDENTIAL****Inventor(s):****Title of the invention:****Manager's Checklist – Please ensure:**

- Clear, readily understandable description of the invention
- Identification of novel features
- Completeness/General presentation
  - Correct forms used
  - Inventorship Statement completed
  - All boxes completed on all forms
  - Forms compiled electronically (optional)

**TO THE MANAGER:**

If you do not consider the subject matter to be suitable for an invention proposal, please seek advice from the Patent Department before signing and forwarding.

**1. Problem addressed or function provided by the invention:** [Example 1a: Finisher cost reduction. 1b: Annotation of copies]

Reduce amount of typing required in producing certain kinds of texts.

**2. New and distinctive feature(s) of the invention:** [Example 2a: New, simplified stacker configuration. 2b: New technique of using low cost LCD to write annotation messages.]

Text is automatically generated from compressed "stenographic" description of its contents.

**3. Could invention have impact beyond current description?** [Example 3a: Could also function for printer finisher. 3b: Could also function to erase edit copy.]

Dictated or manuscript input is also possible.

**4. Potential for Xerox application. Specify product or technology programme if possible.** [Example 4a: Mainline approach in Programme Q. 4b: Adds significant feature to future products.]**5. Value to competitors; potential for license or trade:** [Example 5a: Enables much lower cost finishing than any known system and opens possibilities of moving finishing down-market. 5b: Low cost will be hard to match.]

Lower cost in producing some kinds of texts (because of reduction in typing effort); normalization of output text; higher speed in typing.

**6. Please indicate any related patents, publications, or activities you know of:****Manager:**

I have read and understood the accompanying Invention Proposal, Inventorship Statement Form(s) and above checklist, and agree with the information set out herein.

Signature: Pierre Isabelle

Date: [REDACTED]

**XEROX CONFIDENTIAL****Section 2****Version**

**XEROX****Inventorship Statement Form**

(This form to be *TYPED* and attached to *Invention Proposal Form*  
and *Manager's Comments Form*)

(For Patent Dept. use only)

IP No: 

**XEROX CONFIDENTIAL****Title of the invention:**

**Please explain briefly when and how this proposal was actually devised. If it was devised jointly, explain clearly each individual's contribution to the proposal:**

(Please attach documentary evidence, e.g. extracts from your laboratory notebook(s), technical reports, draft papers or minutes of relevant meetings, wherever possible)

**By signing below, each submitter who claims to be an inventor confirms, that to the best of his/her knowledge, there are no other contributors to the devising of this invention proposal beyond those named herein.**

**If a patent application based on this invention proposal is to be filed, the attorney preparing that application will make the final determination of inventorship.**

*SUBMITTERS/INVENTORS AFTER THE SIXTH, PLEASE USE ADDITIONAL SHEET*

Signed: \_\_\_\_\_  
Full Name: \_\_\_\_\_  
Nationality: \_\_\_\_\_  
Home address: \_\_\_\_\_

Date: \_\_\_\_\_  
Occupation: \_\_\_\_\_  
Location: \_\_\_\_\_

Signed: \_\_\_\_\_  
Full Name: \_\_\_\_\_  
Nationality: \_\_\_\_\_  
Home address: \_\_\_\_\_

Date: \_\_\_\_\_  
Occupation: \_\_\_\_\_  
Location: \_\_\_\_\_

Signed: \_\_\_\_\_  
Full Name: \_\_\_\_\_  
Nationality: \_\_\_\_\_  
Home address: \_\_\_\_\_

Date: \_\_\_\_\_  
Occupation: \_\_\_\_\_  
Location: \_\_\_\_\_

Signed: \_\_\_\_\_  
Full Name: \_\_\_\_\_  
Nationality: \_\_\_\_\_  
Home address: \_\_\_\_\_

Date: \_\_\_\_\_  
Occupation: \_\_\_\_\_  
Location: \_\_\_\_\_

Signed: \_\_\_\_\_  
Full Name: \_\_\_\_\_  
Nationality: \_\_\_\_\_  
Home address: \_\_\_\_\_

Date: \_\_\_\_\_  
Occupation: \_\_\_\_\_  
Location: \_\_\_\_\_

Signed: \_\_\_\_\_  
Full Name: \_\_\_\_\_  
Nationality: \_\_\_\_\_  
Home address: \_\_\_\_\_

Date: \_\_\_\_\_  
Occupation: \_\_\_\_\_  
Location: \_\_\_\_\_

**XEROX CONFIDENTIAL****Section 3****Version:** 

# Semantic Stenography

## 1. The invention: motivation and idea

Professionals who conduct a lot of interviews --- such as sociologists, pollers, doctors, job staffers, ... --- often do not have the time to write down complete well-formed sentences to describe the information communicated to them. Instead it is common for them to jot down as the conversation progresses a few keywords that convey the essential facts. These keywords can later be converted into complete grammatical texts at a more leisurely pace. The ability to perform this fast note-taking task depends on conventions that associate complex conceptual constructions with what could be called "semantic abbreviations". These conventions are sometimes crafted individually ("*idiolect*"), and sometimes shared by communities. The tighter the community and the more frequent the need to communicate recurring types of information, the more efficient such coding can be (this can sometimes go to horrific extremes, as is the case in Simenon's novel "Le Chat", where Emile Bouin and his wife Marguerite have come to hate each other, and continue to communicate only through single content-loaded words scribbled on pieces of paper).

**The present invention** is about automating the process of *converting such short notes*, meaningful inside a restricted community, into a *semantically coherent grammatical text* which is adequate for communication to a wider audience, not privy to the abbreviation conventions used in the restricted community. The method proposed consists in using a document authoring system to model the class of texts in the domain under consideration and in performing a fuzzy match between the given short notes and choices associated with active slots in the authoring system.

The invention is related to methods used in the PhD work of Aurélien Max [M02, MD02], which is oriented towards normalization of legacy documents by inversion of an authoring process. The differences lie (i) in the application to *semantic stenography* and (ii) in the specific techniques that are used in the present case for matching and for user interaction.

## 2. An example

Consider the situation of a large broadcaster of job offer descriptions over the Internet (for instances of such texts, see *Monster - Search Jobs*). One typical job offer for an administrative assistant could read as follows:

*GlobalModest is looking for an Administrative Assistant for its Laval office in France. The position is a CDD for one year to be filled immediately. The main duties will be to schedule appointments, to answer the telephone, to prepare routine letters, to organize and maintain the filing system and to perform a variety of other miscellaneous duties. The candidate should have a Bac+2 level, at least two years experience in a similar position and excellent skills in Word and Outlook. Fluent*

(1)

*knowledge of both French and English are required, with Italian a plus.*

A call-center employee of the broadcaster could take such job offers over the telephone. She could quickly jot down the following notes:

*GlobalModest  
admin assistant  
Laval  
cdd 1 year immediate  
appointments, telephone, simple letters, filing, misc. (2)  
bac+2  
2 years experience  
Word, Outlook  
French, English  
Italian plus*

Then, in a second step, and at a later stage, the broadcaster employee could use these notes to produce the full-text of the job offer, as indicated above.

We will now see one way currently available technology can be applied to semi-automate the second step of this process.


### **3. How the invention works**

The embodiment of the invention that we focus on is based on XRCE's authoring system MDA. This system relies on a formal mechanism (a kind of unification grammar) for describing well-formed semantic representations and their textual realizations in several languages or writing styles. These specifications are restricted to specific domains of discourse for which a relatively complete modelling of document content is possible (such as pharmaceutical leaflets, biological experiment reports, certain types of classified ads, etc.). Such specifications can also be used as enumeration mechanisms which non-deterministically generate the well-formed semantic representations along with their several textual realizations. Authoring then works by asking the user to guide the enumeration process through menu selections associated with different possible paths in the enumeration (for details see [M02].)

An example of a possible semantic structure for the job offer in our example in an MDA system is shown on the left of Figure 1.<sup>1</sup> The textual realization, which will serve as the text of the job offer (1), is shown to the right, with an approximate alignment to elements of the semantic representation.

---

<sup>1</sup> For reasons of exposition, we use a slightly different notation from that actually used for MDA semantic representations.



```

<job_offer>
  <job_description> administrative_assistant </job_description>
  <company> GlobalModest </company>
  <job_location> laval_france </job_location>
  <contract_type> cdd </contract_type>
  <contract_duration> one_year </contract_duration>
  <starting_date> immediate </starting_date>
  <tasks>
    <task> appointments </task>
    <task> telephone </task>
    <task> write_routine_letters </task>
    <task> filing_system_handle </task>
    <task> misc_duties </task>
  </tasks>
  <study_level> bac_plus_2 </study_level>
  <experience_length> 2_years </experience_length>
  <computer_skills>
    <computer_skill>
      <cs_program> Word </cs_program>
      <cs_level> excellent </cs_level>
      <cs_requirement> required </cs_requirement>
    </computer_skill>
    <computer_skill>
      <cs_program> Outlook </cs_program>
      <cs_level> excellent </cs_level>
      <cs_requirement> required </cs_requirement>
    </computer_skill>
  </computer_skills>
  <language_skills>
    <language_skill>
      <ls_idiom> French </ls_idiom>
      <ls_level> fluent </ls_level>
      <ls_requirement> required </ls_requirement>
    </language_skill>
    <language_skill>
      <ls_idiom> English </ls_idiom>
      <ls_level> fluent </ls_level>
      <ls_requirement> required </ls_requirement>
    </language_skill>
    <language_skill>
      <ls_idiom> Italian </ls_idiom>
      <ls_level> good </ls_level>
      <ls_requirement> desirable </ls_requirement>
    </language_skill>
  </language_skills>
</job_offer>

```

GlobalModest is looking for an Administrative Assistant for its Laval office in France.

The position is a CDD for one year to be filled immediately.

The main duties will be to schedule appointments, to answer the telephone, to prepare routine letters, to organize and maintain the filing system and to perform a variety of other miscellaneous duties.

The candidate should have a Bac+2 level, at least two years experience in a similar position and excellent skills in Word and Outlook.

Fluent knowledge of both French and English are required, with Italian a plus.

**Figure 1: Underlying semantic representation and its textual realization (global text).**

The textual realization shown here corresponds to the style of job offer texts (such as can be found on see Monster - Search Jobs), we call it "global text realizations style" to contrast it with another realization style that we present now.

### 3.1. Local realization

Figure 2 is similar to Figure 1, the only difference being that we are now using a different style for realizing the text associated with the semantic representation. We call that the "local text realization style"; it will be handy for two purposes: (i) providing feedback to the user as to the meaning found by the system for different expressions in the short note input, and (2) serving as a basis for the matching procedure between the short notes and possible semantic structures accounting for it. The local text realization is now:

*The job offer is for an administrative assistant.*  
*The hiring company's name is GlobalModest.*  
*The job location is Laval, France.*  
*The contract type is a CDD.*  
*The contract duration is for 1 year.*  
*The position is to be filled immediately.*  
*The job involves handling appointments.*  
*The job involves answering the telephone.*  
*The job involves preparing routine letters.*  
*The job involves handling a filing system.* (3)  
*The job involves other miscellaneous tasks.*  
*At least a bac+2 level is required.*  
*At least 2 years of previous experience are required.*  
*Excellent skills in Word are required.*  
*Excellent skills in Outlook are required.*  
*Fluent knowledge of French is required.*  
*Fluent knowledge of English is required.*  
*Knowledge of Italian would be desirable.*

In Figure 2, the local realization text is now much more closely aligned to the semantic representation than in the global case, often at the level of leaves in the semantic representation; note however that the alignment is sometimes made at the level of small "sub-blocks" in the semantic representation (such as <computer\_skill> or <language\_skill>); these are cases where a finer local realization would not permit a reader to easily reconstruct the scope of the different semantic elements (for instance a realization at the level of the leaves inside the first <computer\_skill> sub-block would lead to the three sentences "Knowledge of Word is desired. The level of knowledge should be excellent. The knowledge of the program is a requirement.")

<pre> &lt;job_offer&gt;   &lt;job_description&gt; administrative_assistant &lt;/job_description&gt;   &lt;company&gt; GlobalModest &lt;/company&gt;   &lt;job_location&gt; laval_france &lt;/job_location&gt;   &lt;contract_type&gt; cdd &lt;/contract_type&gt;   &lt;contract_duration&gt; one_year &lt;/contract_duration&gt;   &lt;starting_date&gt; immediate &lt;/starting_date&gt;   &lt;tasks&gt;     &lt;task&gt; appointments &lt;/task&gt;     &lt;task&gt; telephone &lt;/task&gt;     &lt;task&gt; write_routine_letters &lt;/task&gt;     &lt;task&gt; filing_system_handle &lt;/task&gt;     &lt;task&gt; misc_duties &lt;/task&gt;   &lt;/tasks&gt;   &lt;study_level&gt; bac_plus_2 &lt;/study_level&gt;   &lt;experience_length&gt; 2_years &lt;/experience_length&gt;   &lt;computer_skills&gt;     &lt;computer_skill&gt;       &lt;cs_program&gt; Word &lt;/cs_program&gt;       &lt;cs_level&gt; excellent &lt;/cs_level&gt;       &lt;cs_requirement&gt; required &lt;/cs_requirement&gt;     &lt;/computer_skill&gt;     &lt;computer_skill&gt;       &lt;cs_program&gt; Outlook &lt;/cs_program&gt;       &lt;cs_level&gt; excellent &lt;/cs_level&gt;       &lt;cs_requirement&gt; required &lt;/cs_requirement&gt;     &lt;/computer_skill&gt;   &lt;/computer_skills&gt;   &lt;language_skills&gt;     &lt;language_skill&gt;       &lt;ls_idiom&gt; French &lt;/ls_idiom&gt;       &lt;ls_level&gt; fluent &lt;/ls_level&gt;       &lt;ls_requirement&gt; required &lt;/ls_requirement&gt;     &lt;/language_skill&gt;     &lt;language_skill&gt;       &lt;ls_idiom&gt; English &lt;/ls_idiom&gt;       &lt;ls_level&gt; fluent &lt;/ls_level&gt;       &lt;ls_requirement&gt; required &lt;/ls_requirement&gt;     &lt;/language_skill&gt;     &lt;language_skill&gt;       &lt;ls_idiom&gt; Italian &lt;/ls_idiom&gt;       &lt;ls_level&gt; good &lt;/ls_level&gt;       &lt;ls_requirement&gt; desirable &lt;/ls_requirement&gt;     &lt;/language_skill&gt;   &lt;/language_skills&gt; &lt;/job_offer&gt; </pre>	<p>The job offer is for an administrative assistant  The hiring company's name is GlobalModest  The job location is Laval, France  The contract type is a CDD  The contract duration is for 1 year  The position is to be filled immediately</p> <p>The job involves handling appointments  The job involves answering the telephone  The job involves preparing routine letters  The job involves handling a filing system  The job involves other miscellaneous tasks</p> <p>At least a bac+2 level is required  At least 2 years of previous experience are required</p> <p>Excellent skills in Word are required</p> <p>Excellent skills in Outlook are required</p> <p>Fluent knowledge of French is required</p> <p>Fluent knowledge of English is required</p> <p>Knowledge of Italian would be desirable</p>
---	--

**Figure 2: Local text realization.**

### 3.2. Reconstructing the global text from the short notes

The process of reconstructing the global text (1) from the short notes consists in two steps. The first one consists in producing a fuzzy match between the short notes and the closest semantic structure accounting for them compatible with the MDA specification, the second one a step of realizing the global text corresponding to that structure by using the MDA realization component. This second step is completely standard in MDA and we do not further explain it (see [BDL00]), but not the first step, and we now describe it.

In a nutshell, the matching step on finding matches between on the one hand all local realization statements (a local realization statement being the expressions that appear on a single line of the second column of Figure 2) which are possible relative to the MDA specification and on the other hand subexpressions of the short notes. The best matches are kept and are used to instantiate substructures of the semantic representation.

The matching procedure between a possible local realization statement and short note subexpressions is not purely textual, but can rely on synonymy (simple letter / routine letter; admin / administrative; etc.). Also some words of the local realization statements are given more weight than others, due to their better discriminative power, and carry the main burden of establishing the match (for example in "The contract type is a CDD", the word "CDD" is the heavier word). These aspects are similar to the techniques used in [M02]).

Let's consider an example. The MDA grammar specifies certain combinations of concepts as making sense in the semantic representation. For instance, if we focus on substructures corresponding to *<computer\_skill>*, the MDA grammar specifies which combinations of instances of *<cs\_program>*, *<cs\_level>* and *<cs\_requirement>* are possible. Let's assume that:


*<cs\_program>* is instantiated as a member of the list: Word, Excel, Outlook, ...;  
*<cs\_level>* is instantiated as either: excellent, or some\_experience;  
*<cs\_requirement>* is instantiated as either: required or desirable.

In the second and third cases, one value is considered to be a default and is underlined (but there is no default for the first case).

We assume here that all combinations of instances are accepted by the grammar (this is not always the case, it may be that "some\_experience" would never be "required", but only "desirable", otherwise a stronger level would have been stated.)

Figure 3 shows a partial enumeration of the possible *<computer\_skill>* substructures, with their local realization statements.





<code>&lt;computer_skill&gt;</code>	Excellent skills in Word are required
<code>&lt;cs_program&gt; Word &lt;/cs_program&gt;</code>	
<code>&lt;cs_level&gt; excellent &lt;/cs_level&gt;</code>	
<code>&lt;cs_requirement&gt; required &lt;/cs_requirement&gt;</code>	
<code>&lt;/computer_skill&gt;</code>	
<code>&lt;computer_skill&gt;</code>	Experience with Word is required
<code>&lt;cs_program&gt; Word &lt;/cs_program&gt;</code>	
<code>&lt;cs_level&gt; some_experience &lt;/cs_level&gt;</code>	
<code>&lt;cs_requirement&gt; required &lt;/cs_requirement&gt;</code>	
<code>&lt;/computer_skill&gt;</code>	
<code>&lt;computer_skill&gt;</code>	Excellent skills in Word are desirable
<code>&lt;cs_program&gt; Word &lt;/cs_program&gt;</code>	
<code>&lt;cs_level&gt; excellent &lt;/cs_level&gt;</code>	
<code>&lt;cs_requirement&gt; desirable &lt;/cs_requirement&gt;</code>	
<code>&lt;/computer_skill&gt;</code>	
...	

**Figure 3: Possible values and local realization statements for the `<computer_skill>` substructure.**

We see that the the `<computer_skill>` substructure has a number of possible local realization statements associated with it through the MDA grammar. This is true for the other substructures of Figure 2, and in general the MDA grammar implicitly defines a virtual collection of possible local realization statements.

The matching procedure then works in the following way: it attempts to find virtual local realization statements that "account" for subexpressions of the short notes; during this matching, account is taken of informativeness weights and of synonymy, as sketched above. It may happen that several local realization statements compete for the same short note words; in that case they are ranked according to the tightness of the match; a small premium is given to local realization statement candidates involving default values (such as excellent, or required) to ensure they appear higher on the list than realizations that contain non-default values.

In general, a naive search for a potential local realization statement accounting for some subexpression of a short note could be combinatorially explosive. To avoid this problem in the general case, one can adapt the admissible search procedures described in [M02] for the case of matching a whole *document*. However, our situation here is somewhat simpler: the notion of local realization (as opposed to global realization) makes it realistic to pre-index the substructure types associated with local realizations (such as `<contract_type>`, `<study_level>`, or the slightly more complex `<computer_skill>`) with high-informativity words and their synonyms. This can be used as a first filter permitting to retain only substructure types for which some support exists in the short notes. For each retained substructure type a brute force search for all possible realizations of that type can then be realistically performed to obtain a finer match with the short notes, along with a similarity measure.

In the case of the short notes (2), among the local matches that can be obtained this way we have:

short note:

*admin assistant*

...

*Word*

*Word*

*Word*

*Word*

...

possible match with local realization statement:

The job offer is for an administrative assistant

Excellent skills in Word are required

Experience with Word is required

Excellent skills in Word are desirable

Experience with Word is desirable

Note that the first realization for "Word" is the one which involves the default values "excellent skills" and "required". The ranking would have been different if the short note had been "Word experience a plus".

Figure 4 belows shows a selection of local matches accounting for the intended meaning of the short notes (2). From the local realizations, the semantic structure in the last column can be reconstructed, and from that structure, the global text realization (1) is obtained.

short note	local realization statement	semantic representation
<i>admin assistant</i> <i>GlobalModest</i> <i>Laval</i> <i>cdd</i> <i>1 year</i> <i>immediate</i>	The job offer is for an administrative assistant The hiring company's name is GlobalModest The job location is Laval, France The contract type is a CDD The contract duration is for 1 year The position is to be filled immediately	<pre> &lt;job_offer&gt;   &lt;job_description&gt; administrative_assistant &lt;/job_description&gt;   &lt;company&gt; GlobalModest &lt;/company&gt;   &lt;job_location&gt; laval_france &lt;/job_location&gt;   &lt;contract_type&gt; cdd &lt;/contract_type&gt;   &lt;contract_duration&gt; one_year &lt;/contract_duration&gt;   &lt;starting_date&gt; immediate &lt;/starting_date&gt;   &lt;tasks&gt;     &lt;task&gt; appointments &lt;/task&gt;     &lt;task&gt; telephone &lt;/task&gt;     &lt;task&gt; write_routine_letters &lt;/task&gt;     &lt;task&gt; filing_system_handle &lt;/task&gt;     &lt;task&gt; misc_duties &lt;/task&gt;   &lt;/tasks&gt;   &lt;study_level&gt; bac_plus_2 &lt;/study_level&gt;   &lt;experience_length&gt; 2_years &lt;/experience_length&gt;    &lt;computer_skills&gt;     &lt;computer_skill&gt;       &lt;cs_program&gt; Word &lt;/cs_program&gt;       &lt;cs_level&gt; excellent &lt;/cs_level&gt;       &lt;cs_requirement&gt; required &lt;/cs_requirement&gt;     &lt;/computer_skill&gt;     &lt;computer_skill&gt;       &lt;cs_program&gt; Outlook &lt;/cs_program&gt;       &lt;cs_level&gt; excellent &lt;/cs_level&gt;       &lt;cs_requirement&gt; required &lt;/cs_requirement&gt;     &lt;/computer_skill&gt;   &lt;/computer_skills&gt;   &lt;language_skills&gt;     &lt;language_skill&gt;       &lt;ls_idiom&gt; French &lt;/ls_idiom&gt;       &lt;ls_level&gt; fluent &lt;/ls_level&gt;       &lt;ls_requirement&gt; required &lt;/ls_requirement&gt;     &lt;/language_skill&gt;     &lt;language_skill&gt;       &lt;ls_idiom&gt; English &lt;/ls_idiom&gt;       &lt;ls_level&gt; fluent &lt;/ls_level&gt;       &lt;ls_requirement&gt; required &lt;/ls_requirement&gt;     &lt;/language_skill&gt;     &lt;language_skill&gt;       &lt;ls_idiom&gt; Italian &lt;/ls_idiom&gt;       &lt;ls_level&gt; good &lt;/ls_level&gt;       &lt;ls_requirement&gt; desirable &lt;/ls_requirement&gt;     &lt;/language_skill&gt;   &lt;/language_skills&gt; &lt;/job_offer&gt; </pre>
<i>appointments</i> <i>telephone</i> <i>routine letters</i> <i>filing system</i> <i>misc</i>	The job involves handling appointments The job involves answering the telephone The job involves preparing routine letters The job involves handling a filing system The job involves other miscellaneous tasks	
<i>bac+2</i> <i>2 years</i> <i>experience</i>	At least a bac+2 level is required At least 2 years of previous experience are required	
<i>Word</i>	Excellent skills in Word are required	
<i>Outlook</i>	Excellent skills in Outlook are required	
<i>French</i>	Fluent knowledge of French is required	
<i>English</i>	Fluent knowledge of English is required	
<i>Italian plus</i>	Knowledge of Italian would be desirable	

Figure 4: Matching short notes with local realization statements.

### 3.3. Correcting and post-editing

It may of course happen that the matching procedure that we have described finds the wrong match for some short notes. In such cases, it is convenient to offer the user with the ability to correct some of the matches. One way to do that is to display to her the possible local realization statements aligned with their associated short notes (see Figure 5.) The choice considered as the most probable by the system is displayed first (highlighted here), then other possible choices, and finally the choice "Other" meaning that the user does not accept any of the choices proposed by the system. The figure illustrates three cases where the first system choice is the wrong one, and needs to be corrected by the user (arrow).

The global text realization is shown along with the current state of the selections in the interface, and evolves whenever a different selection is made by the user. For the cases where the user did not agree with any of the system proposals, the corresponding part of the global realization is not shown, and the user has to resort to direct post-editing of the proposed global text.

<i>admin assistant</i>	The job offer is for an administrative assistant Other
<i>GlobalModest</i>	The hiring company's name is GlobalModest Other
<i>Laval</i>	The job location is Laval, Quebec The job location is Laval, France ← Other
<i>cdd</i>	The contract type is a CDD Other
<i>1 year</i>	The contract duration is for 1 year Other
<i>immediate</i>	The position is to be filled immediately Other
<i>appointments</i>	The job involves handling appointments Other
<i>telephone</i>	The job involves answering the telephone Other
<i>simple letters</i>	The job involves preparing routine letters Other
<i>filing system</i>	The job involves creating a filing system The job involves organizing and maintaining the filing system ← Other
<i>misc</i>	The job involves other miscellaneous tasks Other
<i>bac+2</i>	At least a bac+2 level is required Other
<i>2 years experience</i>	At least 2 years of previous experience are required Other
<i>Word</i>	Excellent skills in Word are required Experience with Word is required Excellent skills in Word are desirable Experience with Word is desirable Other
<i>Outlook</i>	Excellent skills in Outlook are required Experience with Outlook is required Excellent skills in Outlook are desirable

	Experience with Outlook is desirable Other
<i>French</i>	Fluent knowledge of French is required Knowledge of French would be desirable Other
<i>English</i>	Fluent knowledge of English is required Knowledge of English would be desirable Other
<i>Italian plus</i>	Fluent knowledge of Italian would be desirable Knowledge of Italian would be desirable ← Fluent knowledge of Italian is required Knowledge of Italian is required Other

Figure 5: A user interface for correcting matching errors.


### 3.4. Other aspects of the invention

**Entity typing.** In the description so far, we have assumed that the MDA grammar was *a priori* aware of all the different possible values for the elements (for instance, the possible values for the type *<cs\_requirement>* are known to be *required* or *desirable*). In fact, certain types, such as *<company>*, have open-ended lists of values that are impossible to specify *a priori*. In intermediary cases such as *<job\_location>*, some values may be known *a priori*, such as important cities, but other values not. In such situations one possible approach is to use entity-typing techniques, such as ThingFinder [T98] or SmartTagging [STXX] to make some guess about the types of the input words that are not known *a priori*.

To give an example, suppose that the short-notes contain the word "Meylan". The entity-type guesser might be able to find as possible type values *<job\_location>* and *<company>*. In this case the user-interface will look like this:

<i>Meylan</i>	The hiring company's name is Meylan The job location is Meylan ← Other
---------------	--

**Context and learning.** It may happen that the expectations about the meaning of a word by the system and by its user are systematically different. For instance, by typing "Outlook", the user might well intend to say that "Experience in Outlook" is required, while the system may get as its first hypothesis that "Excellent skills in Outlook" are required (as in our example in Figure 5). After using the system for a while, the user will come to expect that simply typing "Outlook" later involves a correction of the system choice (by clicking on the second choice proposed, in the example). However, she will quickly learn that by typing "Outlook experience" from the start (or "Outlook exp.", if "exp." is available as a synonym for "experience"), the system will then be able to make the right guess without further interaction. Thus using the system efficiently involves learning to apply the smallest amount of context necessary for "leading" the system to the intended meaning.



This feature of the system is a rather natural and advantageous one, and it is similar to what is common in natural communication, where a speaker tends to unconsciously adapt to the bias of her hearer by providing guiding clues. On the other hand, a better solution would be for the *system* to adapt to the conventions of its user, rather than the reverse. This possibility is outside the scope of this invention, but clearly machine learning techniques could be adapted here in order to learn a re-ranking of the system's proposals on the basis of corrective user clicks.

**Translation.** Another aspect of the invention that is worth mentioning is that (as in Multilingual Document Authoring in general), the approach can be directly extended to the problem of producing a text in another language (say French) than that in which the short notes are written (say English). There is no essential difference between producing French text from the semantic representation obtained from the short notes and producing English text from that representation.